Some notes on biasedness and unbiasedness of two-sample Kolmogorov-Smirnov test

P. Bubeliny
e-mail: bubeliny@karlin.mff.cuni.cz
Charles University, Faculty of Mathematics and Physics, KPMS, Sokolovska 83, Prague, Czech Republic, 18675.

Abstract: *This paper deals with two-sample Kolmogorov-Smirnov test and its biasedness. This test is not unbiased in general in case of different sample sizes. We found out most biased distribution for some values of significance level α. Moreover we discovered that there exists number of observation and significance level α such that this test is unbiased at level α.*

1 Introduction

In the world of statistic, there exists an enormous number of tests and new ones are going to be derived. For most of these tests we know, that they are consistent, we know their asymptotic behavior and a lot of another properties. But there is one thing which is often omitted. This thing is unbiasedness.

Somebody can think, that all of the tests, which are used, are unbiased or are biased against very special alternative which can not occur in practical applications. Somebody can look at unbiasedness as at very poor power of tests against some alternatives and somebody can just thing that unbiasedness is unimportant. But they are all wrong. We often check some assumptions of test by other tests. But what if the checking test is biased and therefore it leads to the bad decision? Then the main test should not be used and it can lead to wrong decision. Therefore, unbiasedness should not be underestimate.

There are a lot of tests which are really unbiased. But there are plenty of tests that are used daily and they are biased. One of such tests is well known two-sample Kolmogorov-Smirnov test. In what follows, we look at biasedness and unbiasedness of this test in some cases in detail.

2 Biasedness and unbiasedness of Kolmogorov-Smirnov test

Firstly, we should recall, what unbiasedness is. A test is said to be unbiased at level α if

1. it has significance level α
2. for all distributions from alternative the power of this test is greater or equal to α.

The test is said to be unbiased if it is unbiased at all level $\alpha \in (0,1)$. Finally, the test is said to be biased if it is not unbiased. Specially, the test is biased at level α against alternative G if it is an level α test and $P(\text{reject}H|G) < \alpha$.

Consider, that $x_1, \dots, x_n$ and $y_1, \dots, y_m$ are two independent samples having distributions with continuous distribution functions F and G, respectively. We would like to test the hypothesis $H : F = G$ against the alternative $A : F \neq G$. Then two-sample Kolmogorov-Smirnov test is based on statistic

$$D_{n,m} = \sup_x |\hat{F}_n(x) - \hat{G}_m(x)|,$$

where $\hat{F}_n(x)$ and $\hat{G}_m(x)$ are empirical distribution functions of F and G. The hypothesis H is rejected for large value of $D_{n,m}$. The exact formula for computing p-values can be found in *Hajek et al.* (1999).

Firstly, we should realize that statistic $D_{n,m}$ of two-sample Kolmogorov-Smirnov test has discrete distribution. Therefore p-values for this test are discrete as well. For example consider that $n = m = 50$. Then the test statistic $D_{n,m}$ can take just 50 different values $1/n, 2/n, \ldots, 1$. For statistic $D_{n,m} = 0.26$ the p-value is equal to 0.0678 and for the next value $D_{n,m} = 0.28$ the p-value is equal to 0.0392. Testing at level $\alpha = 0.05$ could be little bit confusing because the power of this test is equal for each value $\alpha \in [0.0392, 0.0678)$. There exists distribution G such that power of Kolmogorov-Smirnov test at level $\alpha = 0.05$ is equal to 0.045. Such distribution does not meet requirements of definition of unbiasedness for $\alpha = 0.05$ though the power of this test is higher than exact level of this test equal to 0.0392. To hold the idea of unbiasedness for tests with discrete test statistic we should consider just discrete values of significance level α or use randomized versions of these tests.

It should be kept in mind that Kolmogorov-Smirnov test does not depend on monotonic transformation of samples. If we transform both samples (by the same monotonic transformation) to samples with distribution functions F' and G', respectively then $\sup_x |\hat{F}_n(x) - \hat{G}_m(x)| = \sup_x |\hat{F}'_n(x) - \hat{G}'_m(x)|$. Therefore without loss of generality, we assume that F is distribution function of uniform distribution given by

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \le x \le 1 \\ 1 & \text{if } x > 1 \end{cases} . \tag{1}$$

In *Gordon and Klebanov* (2010), they proved that for $n = m$ there exist $\alpha \in (0, 1)$ such that two-sample Kolmogorov-Smirnov test is unbiased at level α against two-sided alternative $F \neq G$. If we consider just one-sided alternatives $A_1 : F \le G$ or $A_2 : F \ge G$ we can extend this founding to $n \neq m$.

Theorem 2.1. *Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$ be independent samples from distribution F and G. Then for arbitrary $n, m \in N$, there exists $\alpha \in (0, 1)$ such that two-sample Kolmogorov-Smirnov test of hypothesis $H : F = G$ against one-sided alternative $A_1 : F \le G$ or $A_2 : F \ge G$ is unbiased at level α.*

Proof. Without loss of generality, we consider that the first sample $x_1, \ldots, x_n$ is from uniform distribution. Firstly, we consider only the alternative $A_1 : F \le G$. For this alternative, the Kolmogorov-Smirnov statistic is given by $D^*_{n,m} = \sup_{x \in (0,1)} \left(\hat{F}_n(x) - \hat{G}_m(x)\right)$, where $\hat{F}_n$ and $\hat{G}_m$ are empirical distribution functions of F and G. The hypothesis H is rejected for small values of $D^*_{n,m}$. Consider α such small, that we reject hypotheses H for $D_{n,m}$ equals to minus one. It occurs if and only if the samples $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$ satisfy

$$\max(y_1, \ldots, y_m) < \min(x_1, \ldots, x_n). \tag{2}$$

The probability of this event is given by

$$n \int_0^1 (1 - x)^{n-1} G^m(x) dx. \tag{3}$$

Moreover, $G(x)$ is monotone and $G(x) \ge x$ because we consider alternative $A_1 : F \le G$. Therefore the function $(1-x)^{n-1}G^m(x)$ of integral (3) attains its minimum for $G(x) = x$. This integral represents probability of rejection of hypothesis at level α if alternative G is true and it is minimized for $F = x = G(x)$. Hence, Kolmogorov-Smirnov test is unbiased at level α.

The proof for alternative $A_2 : F \ge G$ is similar. We take α such small, that we reject hypothesis if and only if $D_{n,m} = 1$. The inequality (2) change to

$$\max(x_1, \ldots, x_n) < \min(y_1, \ldots, y_m)$$

and probability of this event is then given by

$$n \int_0^1 x^{n-1}(1 - G(x))^m dx \tag{4}$$

For alternative A_2, we have $G(x) \le x$ and hence integral (4) is minimized for $G(x) = x$. It proves the theorem. □

The result of this theorem does not mean that two-sample Kolmogorov-Smirnov test is unbiased against one-sided alternative. It only says that there exist small level α for which this test is unbiased. In the following theorem we show that for $n \neq m$ two-sided Kolmogorov-Smirnov test is not unbiased against two-sided alternative.

Theorem 2.2. *Let $x_1, \dots, x_n$ be i.i.d from uniform distribution with distribution function F and $y_1, \dots, y_m$ be i.i.d. from distribution having distribution function G. If $n \neq m$ then there exists $\alpha \in (0, 1)$ such that two-sample Kolmogorov-Smirnov test of hypothesis $H : F = G$ is biased against alternative with the distribution function*

$$G(x) = \frac{(\frac{x}{1-x})^{\frac{n-1}{m-1}}}{1 + (\frac{x}{1-x})^{\frac{n-1}{m-1}}}. \tag{5}$$

Proof. Consider α such small, that we reject hypotheses if and only if $D_{n,m} = \sup_x |\hat{F}_n(x) - \hat{G}_m(x)|$ is equal to one. That is, the samples $x_1, \dots, x_n$ and $y_1, \dots, y_m$ have to satisfy

$$\max(y_1, \dots, y_m) < \min(x_1, \dots, x_n) \text{ or } \max(x_1, \dots, x_n) < \min(y_1, \dots, y_m). \tag{6}$$

The probability of this event is given by

$$n \int_0^1 \Big((1 - x)^{n-1} G^m(x) + x^{n-1}(1 - G(x))^m\Big) dx.$$

Substitute $G(x)$ by y and let the derivative of function $(1 - x)^{n-1} y^m + x^{n-1}(1 - y)^m$ according to y equal to zero. It leads to the equation

$$\Big(\frac{y}{1 - y}\Big)^{m-1} = \Big(\frac{x}{1 - y}\Big)^{n-1}.$$

Therefore the probability of event (6) is not minimized for $F(x) = G(x) = x$ but for

$$G(x) = \frac{(\frac{x}{1-x})^{\frac{n-1}{m-1}}}{1 + (\frac{x}{1-x})^{\frac{n-1}{m-1}}}.$$

□

Some examples of distribution function given by (5) are in figure 1. Although we found out that two-sample Kolmogorov-Smirnov test is biased against alternative (5) it is really true for very small α. Let denote this smallest level α by α_1. Then α_1 can be directly computed by

$$\alpha_1 = n \int_0^1 (1 - x)^{n-1} x^m + x^{n-1}(1 - x)^m \, dx = 2nm \frac{\Gamma(n)\Gamma(m)}{\Gamma(n + m + 1)}. \tag{7}$$

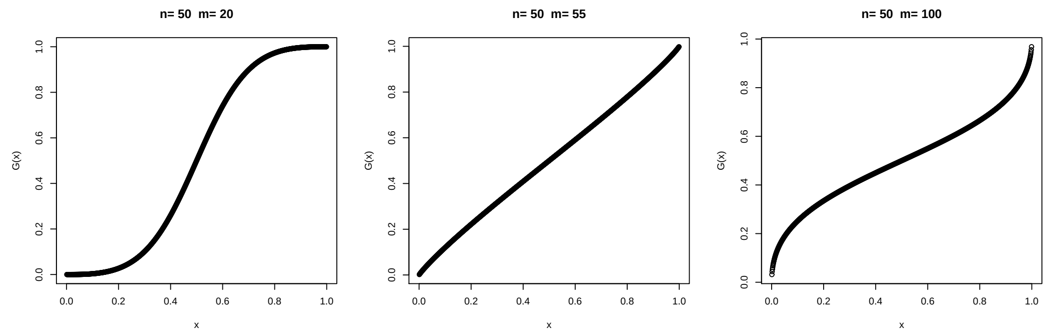

Figure 1: Plot of distribution function G given by (5) for $n = 50$ and $m = 20, 55, 100$

For example if $n = 10$ and $m = 11$ then α_1 is equal to $5.67x10^{-6}$.

All previous result are considered for Kolmogorov-Smirnov statistic $D_{n,m} = 1$. Let consider second highest value of this statistic. For $n > m$ it is equal to $1 - 1/n$ and for $n < m$ it is equal to $1 - 1/m$, respectively. We denote by α_2 the significance level α such that we reject two-sample Kolmogorov-Smirnov test if and only if $D_{n,m} \geq \max(1 - 1/n, 1 - 1/m)$.

Firstly, assume that $n > m \geq 2$ and consider that $D_{n,m} = 1 - 1/n$. It can occur if and only if these samples are such that $x_{(1)} < \ldots < x_{(n-1)} < y_{(1)} < x_{(n)}$ or $x_{(1)} < y_{(m)} < x_{(2)}, \ldots < x_{(n)}$. Together with the case $D_{n,m} = 1$ ($x_{(n)} < y_{(1)}$ or $y_{(m)} < x_{(1)}$) we have that $D_{n,m}$ is greater or equal to $1 - 1/n$ if and only if $x_{(n-1)} < y_{(1)}$ or $y_{(m)} < x_{(2)}$. It leads as to the probability of rejecting the hypotheses at level α_2

$$\begin{aligned} P(D_{n,m} \geq 1 - 1/n) &= P(\forall_j y_j > x_{(n-1)}) + P(\forall_j y_j < x_{(2)}) \\ &= n(n-1) \int_0^1 \Big(x^{n-2}(1-x)(1-G(x))^m + x(1-x)^{n-2} G^m(x) \Big) dx. \end{aligned} \tag{8}$$

As in proof of previous theorem let $G(x) = y$ and let the derivative of interior function of integral (8) according to y equal to zero. It leads us to solve the equation

$$\Big(\frac{y}{1-y}\Big)^{m-1} = \Big(\frac{x}{1-x}\Big)^{n-3}.$$

The solution y as a function of x is given by

$$y = G(x) = \frac{(\frac{x}{1-x})^{\frac{n-3}{m-1}}}{1 + (\frac{x}{1-x})^{\frac{n-3}{m-1}}}. \tag{9}$$

Now assume that $2 \leq n < m$ and consider $D_{n,m} = 1 - 1/m$. It can be true if and only if $y_{(1)} < \ldots < y_{(m-1)} < x_{(1)} < y_{(m)}$ or $y_{(1)} < x_{(n)} < y_{(2)}, \ldots < y_{(m)}$. Therefore the probability of event $D_{n,m} \geq 1 - 1/m$ is equal to

$$P(D_{n,m} \geq 1 - 1/m) \quad = \quad P(D_{n,m} = 1 - 1/m) + P(D_{n,m} = 1)$$

$$= \quad nm\int_0^1 \Big((1-x)^{n-1}G^{m-1}(x)(1-G(x)) + x^{n-1}(1-G(x))^{m-1}G(x)\Big)\,dx$$
$$+n\int_0^1 \Big((1-x)^{n-1}G^{m}(x) + x^{n-1}(1-G(x))^{m}\Big)\,dx. \tag{10}$$

As before let $G(x) = y$ and let the derivative of interior function of integral (10) according to y equal to zero. It leads us to the equation

$$\Big(\frac{y}{1-y}\Big)^{m-3} = \Big(\frac{x}{1-x}\Big)^{n-1}.$$

Therefore the distribution function of most biased distribution in this case is given by

$$y = G(x) = \frac{(\frac{x}{1-x})^{\frac{n-1}{m-3}}}{1+(\frac{x}{1-x})^{\frac{n-1}{m-3}}}. \tag{11}$$

Remark 2.3. *If $n = 3$ and $m = 2$ or $n = 2$ and $m = 3$ then the most biased distribution is discrete distribution given by probabilities $P(y = 0) = P(y = 1) = \frac{1}{2}$ or $P(y = \frac{1}{2}) = 1$, respectively.*

Consider $G(x) = x$ then level α_2 is given (according to (8) and (10)) by

$$\alpha_2 = 2nmk\frac{\Gamma(n)\Gamma(m)}{\Gamma(n+m+1)} = k\alpha_1, \tag{12}$$

where $k = \min(n+1, m+1)$. Distribution functions (9) and (11) are similar to S-curves on figure 1. Although these distribution functions are not equal to themselves and to (5) as well, some interesting results can be found. If $|n-m| = 2$ then (9) and (11) change to $G(x) = x$. It means that the distribution which minimize (8) and (10) is uniform distribution. It leads us to the following theorem.

Theorem 2.4. *Let $\alpha_{n,m}$ be given by (12). If $n = m+2$ or $n = m-2$ then two-sample Kolmogorov-Smirnov test is unbiased at level $\alpha_{n,m}$. Moreover, if $n \neq m$ and $|n-m| \neq 2$ then Kolmogorov-Smirnov test is biased at level $\alpha_{n,m}$.*

Proof. Because of $\alpha_{n,m}$=α_2, the most biased distribution functions are given by (9) and (11). For $|n-m| = 2$ they change to $G(x) = x = F(x)$. It means that the uniform distribution minimize the probability of rejection hypotheses $F = G$ against alternative $F \neq G$ at level α_2 if and only if $|n-m| = 2$. □

Remark 2.5. *If $|n-m| = 1$ then Kolmogorov-Smirnov test is not biased against the distribution functions (9) and (11) at level α_1.*

Let denote by $\mathscr{A}_\alpha$ the set of distributions for which Kolmogorov-Smirnov test is biased at level α, it is

$$\mathscr{A}_\alpha = \{G : P(\text{reject } H \text{ at level } \alpha | \text{alternative } G \text{ is true}) < \alpha\}.$$

For different levels $0 < \alpha < \alpha^*$, one would expect that there is some subset relation between $\mathscr{A}_\alpha$ and $\mathscr{A}_{\alpha^*}$. But it is not generally true. According to the theorem 2.4 there exist G_α such that $G_\alpha \in \mathscr{A}_\alpha$ and $G_\alpha \notin \mathscr{A}_{\alpha^*}$. On the other hand, from remark 2.5 we have that there exists G^*_α such that $G^*_\alpha \notin \mathscr{A}_\alpha$ and $G^*_\alpha \in \mathscr{A}_{\alpha^*}$. Therefore, in general $\mathscr{A}_\alpha$ is not subset of $\mathscr{A}_{\alpha^*}$ and vice versa.

The previous result can be quite simply generalized to α_3 (the third smallest α) in case of $n > 2m$ or $2n < m$. Adding the probability of the even $D_{n,m} = 1 - 2/m$ or $D_{n,m} = 1 - 2/n$ to the (8) or (10) leads us to the most biased distributions at level α_3 given by

$$G_3(x) = \frac{(\frac{x}{1-x})^{\frac{n-5}{m-1}}}{1 + (\frac{x}{1-x})^{\frac{n-5}{m-1}}} \qquad \text{if } n > 2m \tag{13}$$

or

$$G_3(x) = \frac{(\frac{x}{1-x})^{\frac{n-1}{m-5}}}{1 + (\frac{x}{1-x})^{\frac{n-1}{m-5}}} \qquad \text{if } m > 2n. \tag{14}$$

In this case, α_3 is given by

$$\alpha_3 = 2k_2 nm \frac{\Gamma(n)\Gamma(m)}{\Gamma(n+m+1)} = k_2\alpha_1,$$

where $k_2 = \frac{\min((m+2)(m+1),(n+2)(n+1))}{2}$. If $n = m + 4$ or $m = n + 4$ then $G_3(x) = x$. Together with condition $n > 2m$ or $m > 2n$ we have that for $n = 6, m = 2$ or $n = 2, m = 6$ the two-sample Kolmogorov-Smirnov test is unbiased at level $\alpha_3 = 3/7$ and for $n = 7, m = 3$ or $n = 3, m = 7$ the two-sample Kolmogorov-Smirnov test is unbiased at level $\alpha_3 = 1/6$.

Sofar considered α's are too small in case we have some tens of observation in each sample. Therefore we perform the following simulation to look if two-sample Kolmogorov-Smirnov test is biased against the distribution (5) at level $\alpha \approx 0.05$. We set the number of observation n for the first sample be $n = 10, 20, 50, 100$ and the number of observation m for the second sample be $m = 11, 15, 21, 51, 101$. As a distribution of the first sample we consider uniform distribution and for second sample we consider two distributions. The first one is the uniform distribution and the second one is distribution having distribution function G given by (5). We perform 10000 repetitions and compute the difference between the estimate of power if second sample is from alternative distribution and the estimated level α if the second sample is from uniform distribution. The results of this simulation are in table 1. We can see that for all considered n and m the estimate of difference is greater than 0. It means that two-sample Kolmogorov-Smirnov test is not biased against alternative (5) at level $\alpha = 0.05$.

Table 1: Difference between estimate of power for alternative G given by (5) and estimate of level α of two-sample Kolmogorov-Smirnov test.

$\alpha = 5\%$	m=11	m=15	m=21	m=51	m=101
$n = 10$	0.0034	0.0144	0.0320	0.4153	0.7290
$n = 20$	0.0291	0.0087	0.0016	0.2784	0.9170
$n = 50$	0.4071	0.3403	0.2715	0.0001	0.5291
$n = 100$	0.9070	0.9189	0.9190	0.4557	0.0001

3 Conclusion

In this paper we looked at biasedness and unbiasedness of two-sample Kolmogorov-Smirnov test. In case of different sample sizes this test is not unbiased. However we found out that it is not true for all $\alpha \in (0, 1)$.

There exists some special combination of number of observations in each sample and significance level α at which this test is unbiased (see e.g theorem 2.4). Moreover, we discovered the most biased distribution for some values of α. Although we consider just small values of α, for small sample sizes or for data such as gene expressions these levels of α are appropriate. We did not consider all levels of α. However we point out that this test can be unbiased for large samples and α around 0.05. However more research is needed to find out the exact relation between number of observations and level α at which this test is unbiased.

Acknowledgments

The author thanks Prof. Lev Klebanov, DrSc. for valuable comments, remarks and overall help. The work was supported by the grant SVV 261315/2011.

References

Hájek, J., Šidák, Z. and Sen, P., K. (1999), Theory of Rank Tests (Second Edition), *Academic Press*.

Gordon, A., Y. and Klebanov, L., B. (2010), On a paradoxical property of the Kolmogorov-Smirnov two-sample test, *Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in honor of Professor Jana Jurečcková*, Vol. 7, 70-74.

Massey (2010), F., J. (1950), A Note on the Power of a Non-Parametric Test, *Annals of Mathematical Statistics*, Vol. 21, 440-443.